人工智能伦理问题的现状分析与对策

张兆翔[1] 张吉豫[2] 谭铁牛[1*]

1 中国科学院自动化研究所 北京 100190

2 中国人民大学 法学院 北京 100872

摘要 人工智能（AI）是第四次产业革命的核心，但也为伦理道德规范和社会治理带来了挑战。文章在阐释当前人工智能伦理风险的基础上，分析了当前对人工智能伦理准则、治理原则和治理进路的一些共识，提出了以“共建共治共享”为指导理论，逐渐建设形成包含教育改革、伦理规范、技术支撑、法律规制、国际合作在内的多维度伦理治理体系等对策建议。

关键词 人工智能，伦理，治理，应对策略

DOI 10.16418/j.issn.1000-3045.20210604002

人工智能（AI）是第四次产业革命中的核心技术，得到了世界的高度重视。我国也围绕人工智能技术制定了一系列的发展规划和战略，大力推动了我国人工智能领域的发展。然而，人工智能技术在为经济发展与社会进步带来重大发展机遇的同时，也为伦理规范和社会法治带来了深刻挑战。2017 年，国务院印发的《新一代人工智能发展规划》①提出“分三步走”的战略目标，掀起了人工智能新热潮，并明确提出要“加强人工智能相关法律、伦理和社会问题研究，建立保障人工智能健康发展的法律法规和伦理道德框架”。2018 年，习近平总书记在主持中共中央政治局就人工智能发展现状和趋势举行的集体学习时强调，要加强人工智能发展的潜在风险研判和防范，维护人民利益和国家安全，确保人工智能安全、可靠、可控。要整合多学科力量，加强人工智能相关法律、伦理、社会问题研究，建立健全保障人工智能健康发展的法律法规、制度体系、伦理道德。2019 年，我国新一代人工智能发展规划推进办公室专门成立了新一代人工智能治理专业委员会，全面负责开展人工智能治理方面政策体系、法律法规和伦理规范研究和工作推进。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》中专门强调

* 通信作者
资助项目：中国科学院学部科技伦理研究项目（XBKJLL2018001）
修改稿收到日期：2021年11月1日

① 国务院关于印发新一代人工智能发展规划的通知 . 国发〔2017〕35 号 . (2017-07-08)[2021-07-15]. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.

要"探索建立无人驾驶、在线医疗、金融科技、智能配送等监管框架，完善相关法律法规和伦理审查规则"[②]。这些均体现了我国对人工智能伦理及其治理的密切关注程度和积极推进决心，同时也突出了这一问题的重要性。

1 当前人工智能伦理问题

伦理是处理人与人之间关系、人与社会之间关系的道理和秩序规范。人类历史上，重大的科技发展往往带来生产力、生产关系及上层建筑的显著变化，成为划分时代的一项重要标准，也带来对社会伦理的深刻反思。人类社会于20世纪中后期进入信息时代后，信息技术伦理逐渐引起了广泛关注和研究，包括个人信息泄露、信息鸿沟、信息茧房、新型权力结构规制不足等[1]。信息技术的高速变革发展，使得人类社会迅速迈向智能时代，其突出表现在带有认知、预测和决策功能的人工智能算法被日益广泛地应用在社会各个场景之中；前沿信息技术的综合运用，正逐渐发展形成一个万物可互联、万物可计算的新型硬件和数据资源网络，能够提供海量多源异构数据供人工智能算法分析处理；人工智能算法可直接控制物理设备，亦可为个人决策、群体决策乃至国家决策提供辅助支撑；人工智能可以运用于智慧家居、智慧交通、智慧医疗、智慧工厂、智慧农业、智慧金融等众多场景，还可能被用于武器和军事之中。然而，迈向智能时代的过程如此迅速，使得我们在传统的信息技术伦理秩序尚未建立完成的情况下，又迫切需要应对更加富有挑战性的人工智能伦理问题，积极构建智能社会的秩序[2]。

计算机伦理学创始人 Moore[3]将伦理智能体分为4类：① 伦理影响智能体（对社会和环境产生伦理影响）；② 隐式伦理智能体（通过特定软硬件内置安全等隐含的伦理设计）；③ 显示伦理智能体（能根据情势的变化及其对伦理规范的理解采取合理行动）；④ 完全伦理智能体（像人一样具有自由意志并能对各种情况做出伦理决策）。当前人工智能发展尚处在弱人工智能阶段，但也对社会和环境产生了一定的伦理影响。人们正在探索为人工智能内置伦理规则，以及通过伦理推理等使人工智能技术的实现中也包含有对伦理规则的理解。近年来，越来越多的人呼吁要赋予人工智能机器一定的道德主体地位，但机器能否成为完全伦理智能体存在巨大的争议[4]。尽管当前人工智能在一些场景下的功能或行为与人类接近，但实则并不具有"自由意志"。从经典社会规范理论来看，是否能够成为规范意义上的"主体"来承担责任，并不取决于其功能，而是以"自由意志"为核心来构建的。黑格尔的《法哲学原理》即以自由意志为起点展开。因此，当前阶段对人工智能伦理问题的分析和解决路径构建应主要围绕着前3类伦理智能体开展，即将人工智能定性为工具而非主体。

当前阶段，人工智能既承继了之前信息技术的伦理问题，又因为深度学习等一些人工智能算法的不透明性、难解释性、自适应性、运用广泛等特征而具有新的特点，可能在基本人权、社会秩序、国家安全等诸多方面带来一系列伦理风险。例如：① 人工智能系统的缺陷和价值设定问题可能带来公民生命权、健康权的威胁。2018年，Uber自动驾驶汽车在美国亚利桑那州发生的致命事故并非传感器出现故障，而是由于Uber在设计系统时出于对乘客舒适度的考虑，对人工智能算法识别为树叶、塑料袋之类的障碍物做出予以忽略的决定。② 人工智能算法在目标示范、算法歧视、训练数据中的偏失可能带来或扩大社会中的歧视，侵害公民的平等权。③ 人工智能的滥用可能威胁公民隐私权、个人信息权。④ 深度学习等复杂的人

② 中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要. (2021-03-11)[2021-07-15]. http://www.gov.cn/xinwen/2021-03/13/content_5592681.htm.

工智能算法会导致算法黑箱问题，使决策不透明或难以解释，从而影响公民知情权、程序正当及公民监督权。⑤ 信息精准推送、自动化假新闻撰写和智能化定向传播、深度伪造等人工智能技术的滥用和误用可能导致信息茧房、虚假信息泛滥等问题，以及可能影响人们对重要新闻的获取和对公共议题的民主参与度；虚假新闻的精准推送还可能加大影响人们对事实的认识和观点，进而可能煽动民意、操纵商业市场和影响政治及国家政策。剑桥分析公司利用 Facebook 上的数据对用户进行政治偏好分析，并据此进行定向信息推送来影响美国大选，这就是典型实例。⑥ 人工智能算法可能在更不易于被察觉和证明的情况下，利用算法歧视，或通过算法合谋形成横向垄断协议或轴辐协议等方式，破坏市场竞争环境。⑦ 算法决策在社会各领域的运用可能引起权力结构的变化，算法凭借其可以处理海量数据的技术优势和无所不在的信息系统中的嵌入优势，对人们的权益和自由产生显著影响[5]。例如，银行信贷中通过算法进行信用评价将影响公民是否能获得贷款，刑事司法中通过算法进行社会危害性评估将影响是否进行审前羁押等，都是突出的体现。⑧ 人工智能在工作场景中的滥用可能影响劳动者权益，并且人工智能对劳动者的替代可能引发大规模结构性失业的危机，带来劳动权或就业机会方面的风险。⑨ 由于人工智能在社会生产生活的各个环节日益广泛应用，人工智能系统的漏洞、设计缺陷等安全风险，可能引发个人信息等数据泄露、工业生产线停止、交通瘫痪等社会问题，威胁金融安全、社会安全和国家安全等。⑩ 人工智能武器的滥用可能在世界范围内加剧不平等，威胁人类生命与世界和平……

人工智能伦理风险治理具有复杂性，尚未形成完善的理论架构和治理体系。① 人工智能伦理风险的成因具有多元性，包括人工智能算法的目标失范、算法及系统缺陷、受影响主体对人工智能的信任危机、监管机制和工具欠缺、责任机制不完善、受影响主体的防御措施薄弱等[6]。② 人工智能技术和产业应用的飞速发展，难以充分刻画和分析其伦理风险及提供解决方案。这要求我们必须克服传统规范体系的滞后性，而采用“面向未来”的眼光和方法论，对人工智能的设计、研发、应用和使用中的规范框架进行积极思考和构建，并从确立伦理准则等软法开始，引领和规范人工智能研发应用。

关于人工智能的发展，我们既不能盲目乐观，也不能因噎废食，要深刻认识到它可以增加社会福祉的能力。因此，在人类社会步入智能时代之际，必须趁早从宏观上引导人工智能沿着科学的道路前行，对它进行伦理反思，识别其中的伦理风险及其成因，逐步构建科学有效的治理体系，使其更好地发挥积极价值。

2 人工智能伦理准则、治理原则及进路

当前全球人工智能治理还处于初期探索阶段，正从形成人工智能伦理准则的基本共识出发，向可信评估、操作指南、行业标准、政策法规等落地实践逐步深入，并在加快构建人工智能国际治理框架体系[7]。

2.1 伦理准则

近几年来，众多国家、地区、国际和国内组织、企业均纷纷发布了人工智能伦理准则或研究报告。据不完全统计，相关人工智能伦理准则已经超过 40 项。除文化、地区、领域等因素引起的差异之外，可以看到目前的人工智能伦理准则已形成了一定的社会共识。

近年来，中国相关机构和行业组织也非常积极活跃参与其中。例如：2018 年 1 月，中国电子技术标准化研究院发布了《人工智能标准化白皮书（2018 版）》，提出人类利益原则和责任原则作为人工智能伦理的两个基本原则③；2019 年 5 月，《人工智能

③ 中国电子技术标准化研究院编写，国家标准化管理委员会工业二部指导．人工智能标准化白皮书 2018. (2018-01)[2021-07-15]. http://www.cesi.cn/images/editor/20180124/20180124135528742.pdf.

北京共识》发布，针对人工智能的研发、使用、治理3个方面，提出了各个参与方应该遵循的有益于人类命运共同体构建和社会发展的15条原则④；2019年6月，国家新一代人工智能治理专业委员会发布《新一代人工智能治理原则——发展负责任的人工智能》，提出了人工智能发展的8项原则，勾勒出了人工智能治理的框架和行动指南⑤；2019年7月，上海市人工智能产业安全专家咨询委员会发布了《人工智能安全发展上海倡议》⑥；2021年9月，中关村论坛上发布由国家新一代人工智能治理专业委员会制定的《新一代人工智能伦理规范》⑦等。从发布内容上看，所有准则在以人为本、促进创新、保障安全、保护隐私、明晰责任等价值观上取得了高度共识，但仍有待继续加深理论研究和论证，进一步建立共识。

2.2 治理原则

美国、欧洲、日本等国家和地区在大力推动人工智能技术和产业发展的同时，高度重视人工智能的安全、健康发展，并将伦理治理纳入其人工智能战略，体现了发展与伦理安全并重的基本原则。

习近平总书记高度重视科技创新领域的法治建设问题，强调“要积极推进国家安全、科技创新、公共卫生、生物安全、生态文明、防范风险、涉外法治等重要领域立法以良法善治保障新业态新模式健康发展”[8]。近年来，我国在应对新技术新业态的规制和监管方面，形成了“包容审慎”的总体政策。这项基本政策在2017年就已正式提出。在2020年1月1日起实施的《优化营商环境条例》第55条中更是专门规定了“包容审慎”监管原则：“政府及其有关部门应当按照鼓励创新的原则，对新技术、新产业、新业态、新模式等实行包容审慎监管，针对其性质、特点分类制定和实行相应的监管规则和标准，留足发展空间，同时确保质量和安全，不得简单化予以禁止或者不予监管。”⑧这为当前人工智能伦理治理提供了基本原则和方法论。一方面，要注重观察，认识到新技术新事物往往有其积极的社会意义，亦有其发展完善的客观规律，应予以一定空间使其能够发展完善，并在其发展中的必要之处形成规制方法和措施。另一方面，要坚守底线，包括公民权利保护的底线、安全的底线等。对于已经形成高度社会共识、凝结在法律之中的重要权益、价值，在执法、司法过程中都要依法进行保护。这既是法律对相关技术研发者和使用者的明确要求，也是法律对于在智能时代保护公民权益、促进科技向善的郑重承诺。

2.3 治理进路

在人工智能治理整体路径选择方面，主要有两种理论：“对立论”和“系统论”[9]。

*“对立论”主要着眼于人工智能技术与人类权利和福祉之间的对立冲突，进而建立相应的审查和规制制度。*在这一视角下，一些国家和机构重点关注了针对人工智能系统本身及开发应用中的一些伦理原则。例如，2020年《人工智能伦理罗马倡议》中提出7项

④ 北京智源人工智能研究院．人工智能北京共识．(2019-05-25)[2021-07-15]. https://www.baai.ac.cn/news/beijing-ai-principles.html.

⑤ 国家新一代人工智能治理专业委员会．新一代人工智能治理原则——发展负责任的人工智能．(2019-06-17)[2021-07-15]. http://www.most.gov.cn/kjbgz/201906/t20190617_147107.html.

⑥ 上海市人工智能产业安全专家咨询委员会．人工智能安全发展上海倡议．(2019-07-02)[2021-07-15]. http://www.cnr.cn/shanghai/tt/20190702/t20190702_524676955.shtml.

⑦ 国家新一代人工智能治理专业委员会．《新一代人工智能伦理规范》发布．(2021-09-25)[2021-09-27]. http://kw.beijing.gov.cn/art/2021/9/26/art_8910_613576.html.

⑧ 优化营商环境条例，2019年10月8日国务院第66次常务会议通过，2019年10月22日中华人民共和国国务院令第722号发布．自2020年1月1日起施行．(2019-10-22)[2021-07-15]. http://www.gov.cn/zhengce/content/2019-10/23/content_5443963.htm.

主要原则——透明、包容、责任、公正、可靠、安全和隐私⑨，欧盟委员会于2019年《可信赖人工智能的伦理指南》中提出人工智能系统全生命周期应遵守合法性、合伦理性和稳健性3项要求⑩，都体现了这一进路。

“系统论”则强调人工智能技术与人类、其他人工代理、法律、非智能基础设施和社会规范之间的协调互动关系。人工智能伦理涉及一种社会技术系统，该系统在设计时必须注意其不是一项孤立的技术对象，而是需要考虑它将要在怎样的社会组织中运作。我们可以调整的不仅仅是人工智能系统，还有在系统中与之相互作用的其他要素；在了解人工智能运作特点的基础上，可以在整个系统内考虑各个要素如何进行最佳调配治理。当前在一些政策和法规中已有一定“系统论”进路的体现。例如，IEEE（电气与电子工程师协会）发布的《合伦理设计》⑪中提出的8项原则之一即为“资质”（competence），该原则提出系统创建者应明确对操作者的要求，并且操作者应遵守安全有效操作所需的知识和技能的原则，这体现了从对使用者要求的角度来弥补人工智能不足的系统论视角，对智能时代的教育和培训提出了新需求。我国国家新一代人工智能治理专业委员会2019年发布的《新一代人工智能治理原则——发展负责任的人工智能》中，不仅强调了人工智能系统本身应该符合怎样的伦理原则，而且从更系统的角度提出了“治理原则”，即人工智能发展相关各方应遵循的8项原则；除了和谐友好、尊重隐私、安全可控等侧重于人工智能开放和应用的原则外，还专门强调了要“改善管理方式”，“加强人工智能教育及科普，提升弱势群体适应性，努力消除数字鸿沟”，“推动国际组织、政府部门、科研机构、教育机构、企业、社会组织、公众在人工智能发展与治理中的协调互动”等重要原则，体现出包含教育改革、伦理规范、技术支撑、法律规制、国际合作等多维度治理的“系统论”思维和多元共治的思想，提供了更加综合的人工智能治理框架和行动指南。基于人工智能治理的特殊性和复杂性，我国应在习近平总书记提出的“打造共建共治共享的社会治理格局”[10]的指导下，系统性地思考人工智能的治理维度，建设多元共治的人工智能综合治理体系。

3 我国人工智能伦理治理对策

人工智能伦理治理是社会治理的重要组成部分。我国应在“共建共治共享”治理理论的指导下，以“包容审慎”为监管原则，以“系统论”为治理进路，逐渐建设形成多元主体参与、多维度、综合性的治理体系。

3.1 教育改革

教育是人类知识代际传递和能力培养的重要途径。通过国务院、教育部出台的多项措施，以及联合国教科文组织发布的《教育中的人工智能：可持续发展的机遇与挑战》⑫、《人工智能与教育的北京共识》⑬等报告可以看到，国内外均开始重视教育的发展改革在人工智能技术发展和应用中有着不可或缺的作用。为更好地支撑人工智能发展和治理，应从4个

⑨ Rome Call for AI Ethics. (2020-02-28)[2021-07-15]. https://www.romecall.org/wp-content/uploads/2021/02/AI-Rome-Call-x-firma_DEF_DEF_con-firme_.pdf.

⑩ The High-Level Expert Group on Artificial Intelligence. Ethics guidelines for Trustworthy AI. (2019-04-08)[2021-07-15]. https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1.

⑪ The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design. [2021-07-15]. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf.

⑫ United Nations Educational, Scientific and Cultural Organization. Artificial Intelligence for Sustainable Development: Synthesis Report. (2019)[2021-07-15]. https://unesdoc.unesco.org/ark:/48223/pf0000370308.

⑬ United Nations Educational, Scientific and Cultural Organization. Beijing Consensus on Artificial Intelligence and Education. (2019-03-19)[2021-07-15]. https://unesdoc.unesco.org/ark:/48223/pf0000368303/PDF/368303qaa.pdf.multi.page=52.

方面进行完善：① 普及人工智能等前沿技术知识，提高公众认知，使公众理性对待人工智能；② 在科技工作者中加强人工智能伦理教育和职业伦理培训；③ 为劳动者提供持续的终身教育体系，应对人工智能可能引发的失业问题；④ 研究青少年教育变革，打破工业化时代传承下来的知识化教育的局限性，回应人工智能时代对人才的需求。

3.2 伦理规范

我国《新一代人工智能发展规划》中提到，“开展人工智能行为科学和伦理等问题研究，建立伦理道德多层次判断结构及人机协作的伦理框架”。同时，还需制定人工智能产品研发设计人员及日后使用人员的道德规范和行为守则，从源头到下游进行约束和引导。当前有 5 项重点工作可以开展：① 针对人工智能的重点领域，研究细化的伦理准则，形成具有可操作性的规范和建议。② 在宣传教育层面进行适当引导，进一步推动人工智能伦理共识的形成。③ 推动科研机构和企业对人工智能伦理风险的认知和实践。④ 充分发挥国家层面伦理委员会的作用，通过制定国家层面的人工智能伦理准则和推进计划，定期针对新业态、新应用评估伦理风险，以及定期评选人工智能行业最佳实践等多种方式，促进先进伦理风险评估控制经验的推广。⑤ 推动人工智能科研院所和企业建立伦理委员会，领导人工智能伦理风险评估、监控和实时应对，使人工智能伦理考量贯穿在人工智能设计、研发和应用的全流程之中[11]。

3.3 技术支撑

通过改进技术而降低伦理风险，是人工智能伦理治理的重要维度。当前，在科研、市场、法律等驱动下，许多科研机构和企业均开展了联邦学习、隐私计算等活动，以更好地保护个人隐私的技术研发；同时，对加强安全性、可解释性、公平性的人工智能算法，以及数据集异常检测、训练样本评估等技术研究，也提出了很多不同领域的伦理智能体的模型结构。当然，还应完善专利制度，明确算法相关发明的可专利性，进一步激励技术创新，以支撑符合伦理要求的人工智能系统设计[12]。

此外，一些重点领域的推荐性标准制定工作也不容忽视。在人工智能标准制定中，应强化对人工智能伦理准则的贯彻和支撑，注重对隐私保护、安全性、可用性、可解释性、可追溯性、可问责性、评估和监管支撑技术等方面的标准制定，鼓励企业提出和公布自己的企业标准，并积极参与相关国际标准的建立，促进我国相关专利技术纳入国际标准，帮助我国在国际人工智能伦理准则及相关标准制定中提升话语权，并为我国企业在国际竞争中奠定更好的竞争优势。

3.4 法律规制

法律规制层面需要逐步发展数字人权[13]、明晰责任分配、建立监管体系、实现法治与技术治理有机结合。在当前阶段，应积极推动《个人信息保护法》《数据安全法》的有效实施，开展自动驾驶领域的立法工作；并对重点领域的算法监管制度加强研究，区分不同的场景[14]，探讨人工智能伦理风险评估、算法审计、数据集缺陷检测、算法认证等措施适用的必要性和前提条件，为下一步的立法做好理论和制度建议准备。

3.5 国际合作

当前，人类社会正步入智能时代，世界范围内人工智能领域的规则秩序正处于形成期。欧盟聚焦于人工智能价值观进行了许多研究，期望通过立法等方式，将欧洲的人权传统转化为其在人工智能发展中的新优势。美国对人工智能标准也尤为重视，特朗普于 2019 年 2 月发布“美国人工智能计划”行政令，要求白宫科技政策办公室（OSTP）和美国国家标准与技术研究院（NIST）等政府机构制定标准，指导开发可靠、稳健、可信、安全、简洁和可协作的人工智能系统，并呼吁主导国际人工智能标准的制定。

我国在人工智能科技领域处于世界前列，需要更

加积极主动地应对人工智能伦理问题带来的挑战，在人工智能发展中承担相应的伦理责任；积极开展国际交流，参与相关国际管理政策及标准的制定，把握科技发展话语权；在最具代表性和突破性的科技力量中占据发展的制高点，为实现人工智能的全球治理作出积极贡献。

参考文献

1 段伟文. 信息文明的伦理基础. 上海: 上海人民出版社, 2020: 44-94.

2 张文显. 构建智能社会的法律秩序. 东方法学, 2020, (5): 7-9.

3 Moor J. The nature, importance and difficulty of machine ethics. IEEE Intelligence Systems, 2006, 21(4): 18-21.

4 古天龙, 李龙. 伦理智能体及其设计：现状和展望. 计算机学报, 2021, 44(3): 634.

5 张凌寒. 权力之治：人工智能时代的算法规制. 上海: 上海人民出版社, 2021: 34-39.

6 苏宇. 算法规制的谱系. 中国法学, 2020, (3): 166-169.

7 王亦菲, 韩凯峰. 数字经济时代人工智能伦理风险及治理体系研究. 信息通信技术与政策, 2021, (2): 33.

8 习近平. 以科学理论指导全面依法治国各项工作（2020年11月16日）// 论坚持全面依法治国. 北京: 中央文献出版社, 2020: 4.

9 Dubber M, Pasquale F, Das S. The Oxford Handbook of Ethics of AI. Oxford: Oxford University Press, 2020: 48-49.

10 习近平. 决胜全面建成小康社会 夺取新时代中国特色社会主义伟大胜利——在中国共产党第十九次全国代表大会上的报告. 北京: 人民出版社, 2017: 49.

11 郭锐. 人工智能的伦理和治理. 北京: 法律出版社, 2020: 190-195.

12 张吉豫. 人工智能良性创新发展的法制构建思考. 中国法律评论, 2018, 20(2): 116-118.

13 马长山. 智慧社会背景下的“第四代人权”及其保障. 中国法学, 2019, (5): 5-24.

14 丁晓东. 论算法的法律规制. 中国社会科学, 2020, (12): 138-159.

Analysis and Strategy of AI Ethical Problems

ZHANG Zhaoxiang[1] ZHANG Jiyu[2] TAN Tieniu[1*]

（1 Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;

2 Law School, Renmin University of China, Beijing 100872, China）

Abstract Artificial intelligence (AI) is the core of the fourth industrial revolution, and it has brought challenges to ethics and social governance. On the basis of explaining the current ethical risks of artificial intelligence, the study furtherly analyzes the current consensus on ethics, governance principles, and governance approaches of artificial intelligence. Moreover, the study also proposes to take “co-construction, co-governance and sharing” as the guiding theory to gradually build a multi-dimensional ethical governance system, including education reform, ethical norms, technical supports, legal regulations, and international cooperation.

Keywords artificial intelligence (AI), ethics, governance, solution strategies

*Corresponding author



张兆翔　中国科学院自动化研究所研究员。研究方向为生物认知启发的视觉感知与理解、类人学习、类脑智能等。E-mail: zhaoxiang.zhang@ia.ac.cn

ZHANG Zhaoxiang　Professor of the Institute of Automation, Chinese Academy of Sciences (CAS). He specifically focuses on biologically-inspired visual computing, human-like learning, brain-inspired intelligence and their applications on human analysis and scene understanding.
E-mail: zhaoxiang.zhang@ia.ac.cn



谭铁牛　中国科学院院士，英国皇家工程院外籍院士，发展中国家科学院院士和巴西科学院通讯院士。中国科学院自动化研究所智能感知与计算研究中心主任、研究员。研究方向为生物特征识别、图像视频理解和信息内容安全。E-mail: tnt@nlpr.ia.ac.cn

TAN Tieniu　Academician of Chinese Academy of Sciences (CAS), Fellow of the World Academy of Sciences for the advancement of science in developing countries (TWAS), Corresponding Member of Brazil Academy of Sciences, International Fellow of the UK Royal Academy of Engineering, and Professor of Institute of Automation, CAS. He is currently the Director of Center for Research on Intelligent Perception and Computing of Institute of Automation, CAS. His current research interests include biometrics, image and video understanding, and information content security. E-mail: tnt@nlpr.ia.ac.cn

■责任编辑：文彦杰